

Université Paris 7 – Denis Diderot

P.C.E.M. 1

Année universitaire 2007-2008

Travaux Dirigés de Biostatistiques

2. Statistiques

Convention : on considérera pour simplifier les calculs que $1.96 \approx 2$.

NB : les exercices peuvent se faire sans calculatrice sauf ceux devant lesquels il y a le sigle



Propriété de la faculté de Médecine Paris 7 – Denis Diderot

TD4 : estimation, tests d'hypothèses
Cours 6 et 7

Exercice 1

On cherche à estimer la moyenne théorique μ de X à partir d'un échantillon de $2n$ réalisations x_i . Afin d'économiser du temps calcul, un informaticien programme plusieurs formules d'estimation et souhaite vérifier leurs propriétés.

Le premier estimateur utilise seulement les n premières valeurs et a pour formule:

$$T_{2n} = \frac{1}{2n} \sum_{i=1}^n X_i$$

Le deuxième estimateur utilise seulement les n dernières valeurs et a pour formule :

$$U_{2n} = \frac{1}{n} \sum_{i=n+1}^{2n} X_i$$

Q1. Parmi les propositions suivantes la(les)quelle(s) est(sont) vraie(s) ?

- a. T_{2n} est un estimateur non biaisé de μ
- b. U_{2n} est un estimateur non biaisé de μ
- c. La variance de U_{2n} est plus grande que celle de l'estimateur usuel $M_{2n} = \frac{1}{2n} \sum_{i=1}^{2n} X_i$
- d. La loi de U_{2n} est une loi normale
- e. Quand n tend vers l'infini, la loi de U_{2n} tend vers une loi normale

Exercice 2

Soit un échantillon représentatif de 100 femmes primipares. On s'intéresse à la répartition de la durée du travail (en heures) lors de leur accouchement :

Q2. On se propose d'estimer la moyenne m et la variance s^2 de la durée du travail des femmes primipares. On donne $\sum x_i = 700$ et $\sum x_i^2 = 5790$. Parmi les propositions suivantes, la(les)quelle(s) est (sont) vraie(s) ?

- a - m est une estimation ponctuelle de la moyenne
- b - $6 \leq m < 9$ et $2 \leq s^2 < 3$
- c - $7 \leq m < 8$ et $2 \leq s \leq 3$
- d - $7 \leq m < 8$ et $8 \leq s^2 \leq 9$
- e - $8 \leq m < 9$ et $7 \leq s^2 < 9$

Q3. Soit μ la moyenne de la variable aléatoire considérée dans la population et $IC_{95\%}$ son intervalle de confiance à 95%. Parmi les propositions suivantes, laquelle (lesquelles) est (sont) vraie(s) ?

- a - $IC_{95\%}$ est aléatoire
- b - $IC_{95\%}$ n'est pas aléatoire
- c - $IC_{95\%}$ vaut environ $[1,4 ; 12,6]$
- d - $IC_{95\%}$ vaut environ $[6,4 ; 7,6]$
- e - il y a 95 chances sur 100 que m soit compris dans $IC_{95\%}$

Q4. On s'intéresse au pourcentage de femmes primipares dans la population ayant eu une durée de travail strictement supérieure à 10 h. Dans cet échantillon, 14 femmes ont eu un travail de plus de 10 h. On arrondira les proportions à 2 chiffres après la virgule.

Parmi les propositions suivantes, laquelle (lesquelles) est (sont) vraie(s) ?

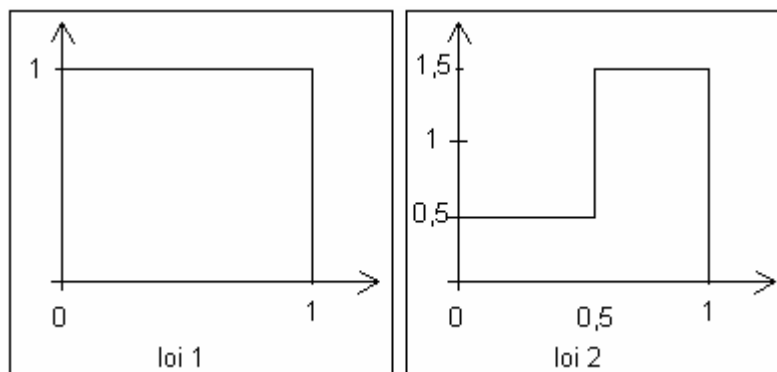
- a - ce pourcentage est égal à 14%
- b - ce pourcentage peut être estimé à 14%
- c - l'intervalle de confiance à 95% de ce pourcentage est environ [7% ; 21%]
- d - l'intervalle de confiance à 95% de ce pourcentage est environ [0% ; 84%]
- e - un intervalle de confiance apporte toujours plus d'information qu'une estimation ponctuelle

Q5. Sur un autre échantillon de taille supérieure à 100, on trouve aussi 14% de femmes primipares dont le travail a une durée supérieure à 10 heures. Parmi les propositions suivantes, laquelle (lesquelles) est (sont) vraie(s) ?

- a - l'intervalle de confiance à 95% calculé sur cet échantillon est plus étendu que celui calculé précédemment
- b - l'intervalle de confiance à 95% calculé sur cet échantillon est moins étendu que celui calculé précédemment
- c - l'intervalle de confiance à 95% est plus étendu que l'intervalle de confiance à 90%
- d - si $n = 200$, les conditions de validité pour le calcul de l'IC_{95%} sont $200 \times 0,14 \geq 5$ et $200 \times (1-0,14) \geq 5$
- e - les conditions de validité du calcul d'un intervalle de confiance d'une proportion sont les mêmes que celles du calcul d'un intervalle de confiance d'une moyenne d'une variable aléatoire continue

Exercice 3

Soit X une variable aléatoire et x une réalisation de X . X peut suivre deux lois dont les densités de probabilité sont données ci-dessous.

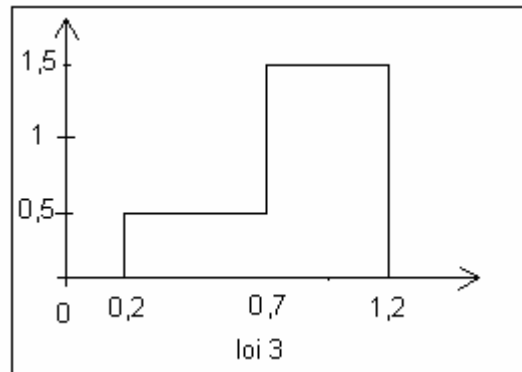


L'hypothèse nulle à tester est que X est distribuée selon la loi 1 contre l'hypothèse alternative que X est distribuée selon la loi 2.

Q6. On définit la région de rejet de H_0 par $\{x \geq 0,5\}$. Parmi les propositions suivantes, laquelle (lesquelles) est (sont) vraie(s) ?

- a - le risque de première espèce est égal à 50%
- b - le risque de première espèce est égal à 25%
- c - la puissance est la probabilité de ne pas rejeter H_0 , à tort
- d - la puissance est égale à 50%
- e - le risque de deuxième espèce est égal à 25%

Q7. L'hypothèse nulle ne change pas, mais l'hypothèse alternative est désormais que X est distribué suivant la loi 3 :



On définit toujours la région de rejet de H_0 par $\{x \geq 0,5\}$. Parmi les propositions suivantes, laquelle (lesquelles) est (sont) vraie(s) ?

- a - le risque de première espèce est égal à 25%
- b - le risque de première espèce est égal à 50%
- c - le risque de deuxième espèce ne change pas
- d - la puissance est égale à 85%
- e - si on avait défini la région de rejet de H_0 par $\{x \leq 0,5\}$, la puissance aurait été meilleure

Exercice 4

Une hypothèse simple consiste à considérer qu'une certaine population est formée pour $\pi = 50\%$ d'hommes et pour $1-\pi = 50\%$ de femmes. Sur un échantillon de 10 000 personnes représentatif de cette population, on trouve une proportion observée p de 48,8% d'hommes.

Q8. Parmi les propositions suivantes, laquelle (lesquelles) est (sont) vraie(s) ?

- a - si la proportion d'hommes dans la population est 50%, l'intervalle de pari à 95% du pourcentage d'hommes dans l'échantillon est environ [47,8% ; 49,8%]
- b - si la proportion d'hommes dans la population est 50%, l'intervalle de pari à 95% du pourcentage d'hommes dans l'échantillon est environ [49% ; 51%]
- c - l'hypothèse nulle à tester est $H_0 : \{p = 0,5\}$
- d - l'hypothèse nulle à tester est $H_0 : \{\pi = 0,5\}$
- e - il faudrait calculer l'intervalle de confiance à 95% du pourcentage d'hommes pour conclure

Q9. Parmi les propositions suivantes, laquelle (lesquelles) est (sont) vraie(s) ?

- a - la région de rejet de l'hypothèse nulle au risque de 5% est $\{p \notin [49,02\% ; 50,98\%]\}$
- b - 49,1% appartient à la région de rejet
- c - 48,5% appartient à la région de rejet
- d - on ne peut rien conclure quant à l'hypothèse « il y a 50% d'hommes dans la population »
- e - on peut conclure qu'il n'y a pas 50% d'hommes dans la population au risque 5%



Exercice 5

On s'intéresse à la proportion d'internes en médecine de l'AP-HP habitants à Paris intra-muros. On constitue par tirage au sort un échantillon d'internes de l'AP-HP. Dans cet échantillon, 850 internes habitent dans Paris intra muros et 920 habitent en périphérie. Soit π la proportion théorique d'internes de l'AP-HP habitants à Paris intra muros. On arrondira les proportions à un chiffre après la virgule.

Q10. Parmi les propositions suivantes laquelle (lesquelles) sont vraie(s) ?

- a - une estimation de π vaut 48%
- b - π se trouve dans l'intervalle de pari à 95% de la proportion observée
- c - l'intervalle de confiance à 95% de π est [45.6% ; 50.4%]
- d - l'intervalle de confiance à 95% de π est [47.6% ; 52.4%]
- e - on ne peut pas, avec ces données, calculer l'intervalle de pari à 95% de la proportion observée



Exercice 1

Q1. On dispose d'un échantillon de 150 individus sur lesquels on mesure une variable aléatoire X . On donne $\sum x = 30\,225$ et $\sum x^2 = 6\,150\,148$. On souhaite tester si la moyenne de X dans la population d'où est tiré l'échantillon est égale à 205. Parmi les propositions suivantes, laquelle (lesquelles) est (sont) vraie(s)?

- a - l'hypothèse nulle est $\mu = 205$
- b - l'écart type estimé de X est égal à environ 401
- c - il faut supposer la normalité de la variable aléatoire X pour conclure
- d - au risque 5%, on conclut que la moyenne de X dans la population d'où est tiré l'échantillon diffère de 205
- e - au risque 5%, on ne peut pas conclure que la moyenne de X dans la population d'où est tiré l'échantillon diffère de 205



Exercice 2

Une étude sur la séro-prévalence (prévalence d'une sérologie positive) du virus HHV8 a été conduite. Parmi les 133 sujets d'origine européenne (E) de l'échantillon étudié, 10 avaient une sérologie positive et parmi les 84 sujets originaires d'Afrique Centrale (AC), 22 avaient une sérologie positive. On se demande si la séro-prévalence (π) est différente entre les 2 populations E et AC.

Q2. Parmi les propositions suivantes, laquelle (lesquelles) est (sont) vraie(s) ?

- a - l'hypothèse nulle, notée H_0 , s'énonce : $p_E = p_{AC}$, où p est la proportion observée
- b - H_0 s'énonce $\pi_E = \pi_{AC}$
- c - l'estimation ponctuelle de la séro-prévalence dans la population E est d'environ 8%
- d - la puissance du test est la probabilité d'accepter H_0 sous H_1 , H_1 étant l'hypothèse alternative
- e - on peut calculer l'intervalle de confiance de p_E

Q3. Soit z le paramètre du test. Parmi les propositions suivantes, laquelle (lesquelles) est (sont) vraie(s) ?

- a - la différence des proportions théoriques suit approximativement une loi normale de moyenne nulle sous H_0
- b - $\{z : |z| < 1,96\}$ est la région de rejet de H_0 (on considérera un risque $\alpha = 5\%$)
- c - quand on fait le test, on trouve $|z| < 1,96$
- d - quand on fait le test, on trouve $|z| > 1,96$
- e - la puissance du test est 90%

Q4. Parmi les propositions suivantes, laquelle (lesquelles) est (sont) vraie(s) ?

- a - on conclut que la séro-prévalence dans l'échantillon E est différente de celle de l'échantillon AC au risque de 5%
- b - on conclut que la séro-prévalence dans l'échantillon E est différente de celle de l'échantillon AC au risque de 1%
- c - on conclut que la séro-prévalence dans la population E est différente de celle de la population AC au risque de 5%
- d - on conclut que la séro-prévalence dans la population E est différente de celle de la population AC au risque de 1%
- e - on ne peut exclure une erreur de première espèce

Exercice 3

L'alfa-L-fucosidase est une enzyme lysosomiale dont l'activité peut être modifiée par certains états pathologiques. La moyenne dans la population générale saine (G) est 4,7 unités. L'alfa-L-fucosidase a été dosée chez 36 sujets atteints de cirrhose éthylique (C) ($m_C = 9,7$ unités, $s_C = 5,0$ unités) et chez 36 sujets diabétiques (D) ($m_D = 6,2$ unités, $s_D = 3,0$ unités).

Q5. On souhaite comparer successivement chacune de ces deux populations (C et D) à la population saine (G). On décide de calculer la statistique de l'écart réduit Z. Parmi les propositions suivantes, laquelle (lesquelles) est (sont) vraie(s) ?

- a - la zone de rejet pour Z ne dépend pas du risque de première espèce
- b - l'alfa-L-fucosidase diffère significativement en moyenne entre les populations saine et cirrhotique, au risque 5%
- c - la probabilité d'observer $|Z| > 5,4$ sous l'hypothèse nulle est inférieure à 0,001 quand on compare les populations saine et cirrhotique
- d - l'ensemble des valeurs de m_D qui conduisent à rejeter l'hypothèse nulle ne dépend pas du nombre de sujets dans l'échantillon
- e - on peut conclure que l'alfa-L-fucosidase diffère significativement entre les patients diabétiques et la population saine, au risque 1%

Q6. On se demande ensuite si l'alfa-L-fucosidase est comparable en moyenne entre les populations cirrhotiques et diabétiques. Parmi les propositions suivantes, laquelle (lesquelles) est (sont) vraie(s) ?

- a - il n'existe pas de différence significative au risque de 5% entre les moyennes des patients cirrhotiques et des patients diabétiques
- b - il existe une différence significative au risque de 5% entre les moyennes des patients cirrhotiques et des patients diabétiques
- c - le degré de signification est $< 0,01$
- d - sous H_0 , la différence entre les moyennes observées chez les patients cirrhotiques et diabétiques est nulle
- e - sous H_0 , la différence entre les moyennes théoriques chez les patients cirrhotiques et diabétiques est nulle

Q7. On suppose que l'hypothèse alternative est que la différence entre les moyennes théoriques des patients diabétiques et des patients cirrhotiques est de 4,0 unités, et on fixe le risque de première espèce à 5%. On supposera que la différence des moyennes a une variance de 1.

Parmi les propositions suivantes, laquelle (lesquelles) est (sont) vraie(s) ?

- a - la puissance du test est comprise entre 80% et 90%
- b - la puissance du test est comprise entre 90% et 95%
- c - la puissance du test est comprise entre 95% et 100%
- d - lorsque la variance du critère étudié augmente, la puissance du test augmente
- e - lorsque la variance du critère étudié diminue, la puissance du test augmente

Q8. Si l'on avait voulu mettre en évidence une différence de 4,0 unités entre les moyennes théoriques des patients cirrhotiques et des patients diabétiques avec une puissance de 99% et un risque de première espèce de 5%, combien de sujets (n) aurait-il fallu inclure par groupe ? On supposera que l'écart type de l'alfa-L-fucosidase vaut 4 unités.

- a - $34 < n < 40$
- b - $39 < n < 45$
- c - $44 < n < 50$
- d - $49 < n < 55$
- e - $54 < n < 60$

Exercice 4

On souhaite étudier l'apport de la radiothérapie locale après la chirurgie dans le traitement du cancer de la prostate non métastasé. On suit les patients pendant un an. La proportion de rechute après chirurgie seule est de 30 %.

NB : On arrondit les proportions à deux chiffres après la virgule (exemple 0,38).

Q3. On traite 100 patients par chirurgie puis par radiothérapie. On observe 20 rechutes. Parmi les propositions suivantes la(les)quelle(s) est (sont) vraie(s) ?

- a- l'estimation ponctuelle de la proportion de rechute après chirurgie suivie de radiothérapie est de 0,30
- b- l'intervalle de confiance à 95 % pour la proportion de rechute après chirurgie seule vaut [0,25 ; 0,35]
- c- l'intervalle de confiance à 95 % pour la proportion de rechute après chirurgie suivie de radiothérapie vaut [0,12 ; 0,28]
- d- le nombre de rechute observé après chirurgie seule sur un échantillon de 100 patients suit une loi binomiale d'espérance 30 et de variance 21
- e- le nombre de rechute observé après chirurgie seule sur un échantillon de 100 patients suit approximativement une loi normale d'espérance 30 et de variance 21



Q4. On décide de faire un essai clinique comparatif pour évaluer l'apport de la radiothérapie. Dans un centre, on constitue deux groupes de 100 patients. Dans le premier groupe, on traite les patients par chirurgie seule et on observe 36 rechutes. Dans le deuxième groupe, on traite les patients par chirurgie suivie de radiothérapie et on observe 18 rechutes. Parmi les propositions suivantes la(les)quelle(s) est (sont) vraie(s) ?

- a- les proportions de rechute dans les deux groupes diffèrent significativement au risque 5 %
- b- les proportions de rechute dans les deux groupes ne diffèrent pas significativement au risque 5 %
- c- les proportions de rechute dans les deux groupes sont significativement identiques au risque 5 %
- d- en prenant 30 % comme proportion de référence, et en voulant montrer une différence de 10 %, il aurait fallu au moins 330 sujets par groupe pour avoir un test de puissance 80 % (et de risque de premier espèce de 5 %)
- e- la puissance du test diminue quand on augmente le nombre de sujet par groupe

Exercice 1

Q1. A propos des essais thérapeutiques randomisés (ETR), quelle(s) est (sont) la (les) proposition(s) vraie(s) ?

- a - le médecin connaît le traitement que recevra le malade lorsqu'il décide de l'inclure dans l'ETR
- b - il est souhaitable que le critère de jugement d'un ETR soit évalué par une personne qui ignore le traitement reçu par le patient
- c - prendre les années de naissance paires ou impaires pourrait être une façon d'attribuer le traitement reçu par le patient dans un ETR
- d - l'administration en double aveugle du traitement a pour objectif d'obtenir des groupes de patients comparables au début de l'essai pour tous les caractères connus ou inconnus autres que le traitement
- e - le tirage au sort garantit le maintien de la comparabilité des groupes en cours d'essai et une égalité dans l'appréciation du critère de jugement

Q2. Concernant l'objectif principal d'un ETR, quelle(s) est(sont) la (les) proposition(s) vraie(s) ?

- a - assurer la comparabilité des groupes au cours du suivi
- b - permettre un nombre égal de patients dans les groupes
- c - pouvoir s'affranchir des questions éthiques
- d - déterminer s'il existe entre les groupes une différence pouvant n'être attribuée qu'au seul traitement
- e - aucune de ces réponses



Exercice 2

Q3. Le mélanome est un cancer cutané dont la gravité dépend de son épaisseur en mm. Les cas de mélanomes sont répartis en fonction de leur épaisseur en mm au moment du diagnostic en 4 stades différents : les stades A ($< 0,75$ mm), B (entre 0,75 et 1,50 mm), C (entre 1,50 et 4,00 mm) et D ($> 4,00$ mm). Dans un pays donné, 20% des cas de mélanomes sont au stade A, 50% au stade B, 25% au stade C et 5% au stade D. Dans un centre médical de ce pays, 280 mélanomes ont été diagnostiqués en une année, 28 de stade A, 126 de stade B, 98 de stade C et 28 de stade D.

On se demande si la distribution des mélanomes vus dans le centre médical est la même que celle du pays.

Parmi les propositions suivantes, laquelle (lesquelles) est (sont) vraie(s) ?

- a - on ne peut pas répondre car les conditions de validité du test ne sont pas remplies
- b - la distribution des mélanomes vus dans le centre médical est la même que celle du pays au risque 5%
- c - la distribution des mélanomes vus dans le centre médical diffère de celle du pays au risque 5%
- d - le test du χ^2 a 1 degré de liberté
- e - on pourrait aussi utiliser un test de l'écart réduit

Exercice 3

Q4. On veut savoir si le temps écoulé depuis la vaccination contre une certaine maladie a une influence sur le degré de gravité de la maladie lorsqu'elle apparaît. Les patients sont divisés en trois catégories selon la gravité de leur maladie – légère (L), moyenne (M), ou grave (G) – et en trois autres selon la durée écoulée depuis leur vaccination – moins de 10 ans (A), entre 10 et 25 ans (B), plus de 25 ans (C). Les résultats portant sur 1574 patients sont les suivants :

	A	B	C	Total
G	1	42	230	273
M	6	114	347	467
L	23	301	510	834
Total	30	457	1087	1574

On donne la valeur du paramètre de test $K = 61,4$

Parmi les propositions suivantes, laquelle (lesquelles) est (sont) vraie(s) ?

- a - sous l'hypothèse nulle, il y a indépendance entre la durée écoulée depuis la vaccination et le degré de gravité de la maladie
- b - on ne peut pas faire le test car dans une des cases on observe un effectif inférieur à 5
- c - on rejette l'hypothèse nulle au risque 1%
- d - le degré de signification du test est inférieur à 10^{-3}
- e - la gravité de la maladie et le délai écoulé depuis la vaccination sont significativement liés au risque 5%



Exercice 4

Q5. On souhaite comparer 2 traitements contre les maux de tête. Le premier traitement est un médicament utilisé de façon habituelle dans cette pathologie, alors que le deuxième traitement est un nouveau médicament. L'attribution du traitement s'est faite par tirage au sort. Le traitement est dit efficace si la douleur disparaît dans les 2 heures suivant la prise du médicament. On a obtenu les résultats suivants :

	Patients avec traitement 1	Patients avec traitement 2
Efficace	60	100
Non efficace	140	120

Parmi les propositions suivantes, laquelle (lesquelles) est (sont) vraie(s) ?

- a - on a effectué un ETR (Essai Thérapeutique Randomisé)
- b - on peut utiliser un test du χ^2 à 2 d.d.l.
- c - on rejette l'hypothèse nulle au risque 5%
- d - on peut conclure quant au rejet ou au non rejet de H_0 car une procédure en double aveugle a été utilisée pour l'attribution du traitement
- e - on peut conclure que le traitement 2 est plus efficace que le traitement 1 au risque 5%

Q6. On veut mettre en évidence une différence de 10% entre les pourcentages théoriques de réussite chez les patients sous traitement 1 et chez les patients sous traitement 2, avec une puissance de 99% et un risque α de 5%. On prend comme valeur de π le pourcentage de réussite avec le traitement de référence (traitement 1). Soit n , le nombre de sujets qu'il faut inclure par groupe.

Parmi les propositions suivantes, laquelle (lesquelles) est (sont) vraie(s) ?

- a - sous H_1 , la variable aléatoire différences des proportions observées de réussite sous les traitements 1 et 2 a une variance de 1
- b - pour $\beta = 1\%$, x_β vaut 2,40
- c - $850 < n < 875$
- d - $775 < n < 800$
- e - $750 < n < 775$

Exercice 5

On veut comparer la prévalence d'une maladie M dans deux groupes indépendants de sujets.

Q7. Concernant le nombre de sujets nécessaire à inclure dans chaque groupe, indiquez parmi les propositions suivantes, laquelle (lesquelles) est (sont) vraie(s) ?

- a – il augmente quand la puissance et le risque α augmentent
- b – il augmente quand la puissance augmente et diminue quand le risque α augmente
- c – il augmente quand la différence à mettre en évidence entre les deux groupes augmente
- d – il diminue quand la différence à mettre en évidence entre les deux groupes augmente
- e – il ne peut être calculé s'il ne s'agit pas d'un essai thérapeutique randomisé

Q8. On réalise les observations suivantes :

Groupe	A	B
effectifs	100	300
malades	25	55

Parmi les propositions suivantes, laquelle (lesquelles) est (sont) vraie(s) (au risque 5%) ?

- a - la prévalence de M diffère entre les deux groupes
- b - la prévalence de M ne diffère pas entre les deux groupes
- c - on a démontré que la maladie M est liée au groupe
- d - on a démontré que la maladie M est indépendante du groupe
- e - le degré de signification est inférieur à 0,05